Molecular Nutrition & Food Research
www.mnf-journal.com

A Standardized Strategy for Simultaneous Quantification of Urine Metabolites to Validate Development of a Biomarker Panel Allowing Comprehensive Assessment of Dietary Exposure

*Manfred Beckmann, Thomas Wilson, Hassan Zubair, Amanda J. Lloyd, Laura Lyons, Helen Phillips, Kathleen Tailliart, Nicholas Gregory, Rhys Thatcher, Isabel Garcia-Perez, Gary Frost, John M. Mathers, and John Draper**

Scope: Metabolites derived from individual foods found in human biofluids after consumption could provide objective measures of dietary intake. For comprehensive dietary assessment, quantification methods would need to manage the structurally diverse mixture of target metabolites present at wide concentration ranges.
Methods and results: A strategy for selection of candidate dietary exposure biomarkers is developed. An analytical method for 62 food biomarkers is validated by extensive analysis of chromatographic and ionization behavior characteristics using triple quadrupole mass spectrometry. Urine samples from two food intervention studies are used: a controlled, inpatient study ($n = 19$) and a free-living study where individuals ($n = 15$) are provided with food as a series of menu plans. As proof-of-principle, it is demonstrated that the biomarker panel could discriminate between menu plans by detecting distinctive changes in the concentration in urine of targeted metabolites. Quantitative relationships between four biomarker concentrations in urine and dietary intake are shown.
Conclusion: Design concepts for an analytical strategy are demonstrated, allowing simultaneous quantification of a comprehensive panel of chemically-diverse biomarkers of a wide range of commonly-consumed foods. It is proposed that integration of self-reported dietary recording tools with biomarker approaches will provide more robust assessment of dietary exposure.

1. Introduction

Recent interest in the discovery of objective biomarkers of dietary intake[1] has been stimulated by the observation that metabolites derived from individual foods are present in post-prandial blood and urine samples.[2–5] If validated, such dietary exposure biomarker data could be used to cross-reference and to complement information derived from traditional, self-reported dietary assessment instruments. Importantly, it could help to reduce the bias which is both pervasive and hard to quantify in evaluation of habitual dietary intake. The ideal biomarker is highly specific to one food item or food group, it is not detected in the biological sample of interest when the specific food item is not ingested, and shows distinct dose- and time- dependent responses after consumption of the particular food.[6] Although some metabolites are associated with exposure to a defined food or food group, other putative biomarkers are much less specific and considerable rigor has to be applied in any future validation of their utility

Dr. M. Beckmann, Dr. T. Wilson, Dr. H. Zubair, Dr. A. J. Lloyd, Dr. L. Lyons, H. Phillips, K. Tailliart, Dr. N. Gregory, Dr. R. Thatcher, Prof. J. Draper
Institute of Biological
Environmental and Rural Sciences
Aberystwyth University
Aberystwyth SY23 3DA, UK
E-mail: jhd@aber.ac.uk

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/mnfr.202000517

DOI: 10.1002/mnfr.202000517

Prof. J. M. Mathers
Human Nutrition Research Centre
Population Health Sciences Institute
William Leech Building
Newcastle University
Newcastle upon Tyne NE2 4HH UK

Dr. I. Garcia-Perez, Prof. G. Frost
Nutrition and Dietetic Research Group
Division of Diabetes
Endocrinology and Metabolism
Department of Medicine
Hammersmith Hospital Campus
Imperial College London
London W12 0NN UK

in monitoring dietary intake.[7] For application in the real world, the key need is to monitor simultaneously as wide a range of biomarkers as possible to determine the degree of overlap between foods during any validation process.[8] The issue is complicated because many food-derived chemicals undergo various modifications during digestion and absorption, or in subsequent tissue (liver) metabolism, so that the chemical nature of biomarkers in body fluids or urine differs from those in the ingested foods.[9] In human studies, urine is an easily accessible biofluid that contains relatively high concentrations of many dietary metabolites or their bio-transformation products.[10] Whilst blood fractions may be a preferred source of lipid-related biomarkers (e.g., fatty acids from dairy products and carotenoids from vegetables), urine contains a higher abundance or provides better discrimination for most of the investigated compounds.[11] A literature search initiated in 2010 revealed that > 1000 individual metabolites had been proposed as putative dietary exposure biomarkers (see Table S1, Supporting Information) in urine, plasma, and serum. The water-soluble, putative dietary exposure biomarkers found in urine are derived from a diverse range of chemical classes found in foods, including di-peptides, simple amines, aromatic alcohols, carboxylic acids, carnitines, alkaloids, phenolic acids and are amplified by a wide range of compounds derived from the microbial catabolism of dietary polyphenols.[9] Any analytical method that is intended to be used to support comprehensive dietary exposure assessments should allow simultaneous quantification of the structurally diverse mixture of target metabolites present at a wide range of concentrations in urine. The advent of ultra-high-performance liquid chromatography (UHPLC) columns coupled to highly selective and sensitive triple quadrupole mass spectrometry (QQQ-MS) instruments offers perhaps the best opportunity to develop quantitation approaches suitable for dietary intake surveys.[12,13] To be useful for research or for surveying links between nutrition and health, a panel of dietary biomarkers should be as comprehensive as possible and include all the key foods and food groups that are consumed habitually by the study population.[7,14] In two recent reports, we have described the design of a food intervention study to validate biomarkers performance in the context of the typical UK diet and that included foods processed and prepared in a range of food formulations.[15,16] Key practical aspects of study design relate to the selection of target foods/food groups where biomarker discovery is feasible and relevant to national policies[15] and the development of urine sampling methods, suitable for large scale use in community settings.[16] In the present study we describe progress toward the development of a rapid, robust, and extendable, analytical strategy allowing simultaneous quantification of a comprehensive panel of chemically diverse potential dietary exposure biomarkers for use in future validation studies.

2. Participants and Methods

2.1. Ethical Approval

The metabolomics at Aberystwyth, Imperial, and Newcastle (MAIN) project included two major food interventions. In a randomized, controlled, crossover trial (Study 1), participants were recruited from a database of healthy volunteers at the UK National Institute for Health Research (NIHR)/Wellcome Trust Imperial Clinical Research Facility (CRF). This study was approved by the London–Brent Research Ethics Committee (13/LO/0078). For Study 2, a favorable ethical opinion was obtained following Proportionate Review by the East Midlands–Nottingham 1 National Research Ethics Committee (14/EM/0040),[15,16] and Caldicott approval for storage of data and data protection was granted by Newcastle-upon-Tyne Hospitals NHS Foundation Trust [6896(3109)]. Study 2 was adopted into the UK clinical research network portfolio (16 037) and is registered with international standard randomized controlled trials number, 88 921 234. All participants in both studies gave written informed consent, and studies were carried out in accordance with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Study 3 samples were obtained from a pilot randomized controlled trial (RCT) for evaluating the effectiveness of a one-to-one consultation with a GP Practice Nurse to improve patient lifestyle and reduce the risk of Type-2 diabetes. The trial was conducted by the Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University. A favourable ethical opinion was obtained from the Wales Research Ethical Committee (IRAS 239 910) following proportionate review.

2.2. Study Participants and Food Intervention Studies

The first intervention (Study 1) was designed specifically to provide foods relevant to the World Health Organization (WHO) healthy eating guidelines[17] at different exposure levels in a controlled, inpatient study.[18] The second intervention study (Study 2)[15,16] involved free-living individuals who were given all their foods and beverages to prepare and consume at home using fixed menus that included foods for which putative dietary exposure biomarkers had been described and for which biomarker measurement was deemed feasible.[15,16] For the controlled, inpatient study (Study 1) cumulative and spot urine samples were obtained from 19 healthy individuals aged 21–65 years and mean BMI 25.6 (±3.2 SD kg m^{-2}). Study participants were exposed to menus providing 25%, 50%, 75%, or 100% of the dietary intake recommended by the WHO healthy eating guidelines.[17] The same menu was consumed each day during the 3 days within a single treatment period to ensure a stable dietary exposure. Spot and cumulative urines were collected,[18] including a post-dinner (PD) sample, 2–3 h after the main evening meal, using single-use urine containers (International Scientific Supplies Ltd., Bradford, United Kingdom). Urine samples were stored at −80 °C until analysis.

Study 2 was designed to validate putative exposure biomarkers for commonly consumed foods, typical of UK diets. In two cohorts of participants, foods were consumed as a series of menus on 6 different days using a range of food formulation, processing, and cooking methods and participants collected spot urine samples at home.[15,16] In the second cohort, 15 healthy individuals (aged 21–74 years, mean BMI 24.4 (±2.3 SD kg m^{-2})) took part in the food interventions on 3 different days. Participants were provided with all the food and ingredients to prepare and consumed meals at home following a series of menus designed to generate distinctive “Menu Days”. Participants collected urine samples at pre-determined time points[15,16] including the first morning void (FMV) after each menu day. Urine samples were collected in a plastic jug and a sub-sample was added to a labelled sterile 30 mL Universal tube and placed in an opaque cool bag in the fridge at 4 °C. The rest of the urine was discarded. At the end of the dietary intervention period, all urine samples were brought to the laboratory in a cool bag. Aliquots (2 mL Eppendorfs) of all urine samples were made and stored at −80 °C until analysis.

To assess the typical ranges of urinary biomarkers in a free-living population, multiple FMV urine samples were collected from individuals ($n = 23$, aged 45–78 years, BMI range 19.8–46.7 kg m^{-2}) participating in an ongoing clinical trial (Study 3) who collected three FMV urine samples on non-consecutive days for each of 3 consecutive weeks (total 9 urine samples). Urine samples for each participant in Study 3 were normalized by refractive index (see below) and then pooled, to create a single sample that reflected each individual's habitual diet over a 3-week period. These samples were used to obtain the typical concentration range of individual biomarkers in real world (i.e., non-dietary intervention) conditions.

2.3. Strategy for Selection of Candidate Dietary Exposure Biomarkers

The selection of biomarkers was initiated with a literature search to generate an initial "long list" of food-related metabolites with potential for inclusion in a panel of biomarkers that would provide comprehensive coverage of food items consumed in Studies 1 and 2. The search was carried out using Google Scholar and Web of Knowledge using the following search terms in a range of combinations 'biomarkers + blood + urine + food + dietary. Publications were screened and information was added to the database if they contained data relating to potential dietary exposure biomarkers measured in either blood (mainly plasma) or urine samples (see Table S1, Supporting Information). Examination of the database revealed several major trends. First, of 1153 reported candidate biomarker signals, 976 were discovered in urine and of 155 putative food intake biomarkers reported to be present in plasma/serum, 39 were also found in urine. Based on these data, we focused on urine as a sample type for development of a food intake biomarker panel.

More detailed analysis of the food-related metabolite database revealed that >42% of urinary metabolite signals with dietary exposure biomarker potential were phase II biotransformation products including 120 glucuronides/di-glucuronides, 108 sulfates, and 15 mixed glucuronide/sulfates (Table S1, Supporting Information). With few exceptions, chemical standards for these conjugated metabolites are not available routinely. For these reasons, we restricted the initial list of biomarker candidates (see Table 2) to metabolites for which chemical standards were readily available at reasonable cost (<$100 per mg) and which were stable in calibration stock solutions. Although a few publications reported the use of nuclear magnetic resonance or GC-MS technologies, the majority (>80%) of biomarker signals were quantified using an LC-MS-based method. Based on these data we focused our analytical approach on the use of LC-MS with a standard range triple quadrupole (QQQ) mass spectrometer to investigate the applicability of these methodologies.

2.4. Processing of Urine Samples

Urine samples stored at −80 °C were thawed at room temperature. Samples were vortexed and 800 µL of each sample was transferred to a clean 2 mL Eppendorf tube. Samples were centrifuged (EBA 12R, Hettich Zentrifugen) at 14 000 rpm for 5 min at 4 °C, and 200 µL of supernatant was aliquoted onto a refractometer (OptiDuo 38–53, Bellingham and Stanley) prism surface for three consecutive specific gravity readings. The refractometer prism surface was rinsed with water and dried with tissue between measurements. For urine extraction based on specific gravity correction, urine samples were aliquoted according to specific gravity correction factors (see below) and remaining volume of ultra-pure (18.2 Ω) H_2O was added to make a total volume of 500 µL before adding 500 µL of methanol.[19] Specific gravity correction factors were calculated for participant urine samples as a fold change of each urine specific gravity to the lowest urinary specific gravity measured for that participant. The Eppendorf tube containing urine and methanol was vortexed and then placed on an orbital shaker (FATSM002, Favorgen Biotech Corp) for 20 min at 4 °C and 1400 rpm in the dark. Extracted samples were stored at −80 °C until further analysis.

2.5. Chemicals and Reagents

The following solvents were used for LC mobile phase and sample preparation: Acetonitrile (Optima LC/MS grade, Fisher Scientific, UK) and methanol (HPLC grade for LC analysis, primer trace analysis grade for urine extraction and standards preparation, Fisher Scientific, UK); ammonium acetate (Optima LC/MS grade, Fisher Scientific, Belgium); formic acid (FA, for mass spectrometry, ≈98%, Fluka, Sigma-Aldrich, USA). Water was ultra-pure water (18.2 Ω) drawn from an Elga Purelab flex water purifier system (Taiwan). All chemical standards covering the biomarker panel had a minimal purity of 98%.

2.6. Sample Analysis by Liquid Chromatography Triple Quadrupole Mass Spectrometry

Sample analysis was performed on a TSQ Quantum Ultra EMR QQQ mass spectrometer (Thermo Scientific) equipped with a heated electrospray ionization source. Samples were delivered using an Accela UHPLC system (Thermo Scientific) consisting of autosampler, column heater, and quaternary UHPLC-pump. For hydrophilic interaction liquid chromatography (HILIC) analysis, chromatographic separation was performed on a ZIC-pHILIC (polymeric 5 µm, 150 × 4.6 mm) column (Merck). The mobile phase consisted of 10 mM ammonium acetate in water: acetonitrile (95:5) (A) and 10 mM ammonium acetate in water: Acetonitrile (5:95) (B). The gradient program used was as follows: 0 min, 95% B (400 µL min^{-1}); 15 min, 20% B (400 µL min^{-1}); 15.01 min, 20% B (500 µL min^{-1}); 20 min, 20% B (500 µL min^{-1}); 20.01 min, 95% B (500 µL min^{-1}); 25 min, 95% B (500 µL min^{-1}). The HPLC was carried out in low pressure (≈0–7000 psi) operating mode with 0 and 650 psi as minimum and maximum pressures, respectively.

For reverse phase (RP) analysis, chromatographic separation was performed on a Hypersil Gold (1.9 µm, 200 × 2.1 mm)

RP- column (Thermo Scientific). The mobile phase consisted of 0.1% formic acid in H_2O (**A**) and 0.1% formic acid in MeOH (B). The gradient program used was as follows: 0 min, 0% B; 0.5 min, 0% B; 5 min, 60% B; 11 min, 100% B; 13 min, 100% B; 13.01 min, 0% B; 19 min, 0% B. For RP analysis, the flow rate was maintained at 400 μL min^{-1}. The UHPLC was carried out in high pressure (≈7000–15 000 psi) operating mode with 0 and 1000 psi as minimum and maximum pressures, respectively.

For both chromatographic analyses, column oven and autosampler tray were maintained at 60 and 14 °C, respectively. To ensure consistent sample delivery, 20 μL were injected using a 20 μL loop and partial loop injection mode. After each injection, syringe and injector were cleaned using a 10% HPLC grade MeOH solution in ultra-pure water (1 mL flush volume; 100 μL s^{-1}flush speed, 2 mL wash volume) to avoid sample carryover. Mass spectra were acquired in multiple reaction monitoring (MRM) mode, in positive and negative ionization polarities simultaneously using optimized values of collision energy and tube lens for each MRM transition (Table S2, Supporting Information). Spectra were collected at a scan speed of 0.010 and 0.003 s for HILIC and RP analysis, respectively. A scan width of 0.010 m/z units and peak width (Q1, Q3) of 0.7 FWHM were used for both HILIC and RP analyses.

Raw files (ThermoFisher) were converted to mzML[20] using msconvert in the ProteoWizard tool kit.[21] All further processing of mzML files was performed using the R statistical programming language.[22] Selected reaction monitoring (SRM) chromatograms were extracted from mzML files using the R library, mzR and peaks areas were calculated by extracting pre-defined chromatographic windows (based on calibration standards) around each peak apex. Absolute concentrations were calculated using a nine-point calibration curve (0.006561 to 100 μg mL^{-1}). For each calibration standard a quadratic equation was used to model the relationship between peak area and concentration (Table S2, Supporting Information). A squared fit of log 10-transformed values accommodated best the wide concentration range for biomarkers in high and low consumers of target foods, without compromising accuracy and normal distribution requirements for regression analysis.

The limit of detection (LOD) and limit of quantification (LOQ) of all chemical standards were calculated as the lowest concentration of each biomarker giving a signal-to-noise ratio of 3:1 and 10:1, respectively within the linear range of each calibration curve.

2.7. Quality Control Strategy for Target Biomarkers

Reproducibility of the mixture of chemical standards was determined using the relative standard deviation (RSD) of a multi component calibration standard and an external urine QC sample using a 'master mix' of pooled samples derived from Study 3. The external urine QC sample was used in order to determine the effect of the resultant urine matrix on the reproducibility of selected biomarkers across multiple experiments. The external QC (as opposed to an experimental QC) allowed for longitudinal monitoring of RSD without intra experimental bias.

2.8. Data Analysis

Supervised classification of quantitative metabolite data was performed using random forest (RF) classification using the randomForest package[23] in R.[22] For all RF models, the number of trees (ntree) used was 1000 and the number of variables considered at each internal node (mtry) was the square root of the total number of variables. Accuracy, margins of classification and area under the receiver operator characteristic (ROC) curve (AUC) were all used to evaluate the performance of classification models, as described previously.[24] RF classification models were plotted following multi-dimensional scaling (MDS). Proximity measures for each individual observation were extracted from RF models and scaled coordinates produced using cmdscale on 1—proximity. RF selection frequencies were used as the criteria for determining the explanatory power of individual features using the methods described by.[25] Selection frequency false positive rates were calculated using the R package forestControl (available at http:/github.com/wilsontom/forestControl).

3. Results

3.1. Establishment of a Food Biomarker Selection and Analytical Strategy

Table 1 summarizes the number of metabolites suggested in current literature (Table S1, Supporting Information) to be associated with exposure to 49 specific foods/food groups used in the MAIN study (Studies 1 and 2). Urine biomarker leads have been discovered for the vast majority of food commonly consumed in the UK with the exception of leafy green vegetables other than crucifers (e.g., lettuce and spinach), carrots, mushrooms, rice, eggs, and dairy products in general, with the exception of cheese (Table 1**)**. It is clear that there is considerable choice in terms of potential biomarkers, with 25 foods/food groups reported to be represented by 15 or more putative biomarkers.

Our initial LC-MS analytical methods were derived from examination of protocols compiled from the literature search for candidate dietary exposure biomarkers and adapted for targeted, quantitative measurement on a standard range QQQ-MS. As a prelude to examination of chromatographic behavior in mixtures, the mass transition parameters for 62 selected putative food exposure biomarkers were determined individually. For each metabolite, the ionization mode and SRM parameters of collision energy and product ion mass were investigated to optimize the signal-to-noise ratio while also attempting to differentiate isomeric or isobaric compounds by product ion mass (Table S2, Supporting Information). Preliminary assessment of the chromatographic separation characteristics of chemical standards mixtures on several types of UHPLC columns revealed early elution and poor resolution of several, more polar, metabolites with potential as food exposure biomarkers. Such compounds included sucrose, tartarate, 1- and 3-methylhistidine and the dipeptides anserine and carnosine which exhibited excellent chromatographic characteristics (**Figure 1**A,B) when analyzed on HILIC columns.[26] To ensure inclusion of such metabolites in a comprehensive biomarker detection and quantitation panel, we adopted the use of both RP and HILIC columns in all subsequent experiments. Extensive analysis of chromatographic behavior characteristics

Table 1. Summary of putative urinary biomarkers for 49 foods and food groups derived from published literature.

Dietary components targeted in MAIN study menus	No. of putative biomarkers	Main biofluid type	Main chemical classes
Beverages			
Tea	131	Urine	Polyphenol-derived conjugates and gallates
Coffee	74	Urine	Phenolic acids, alkaloids, terpenes
Confectionary/sweetened drinks			
Cocoa products/chocolate	73	Urine	Methylxanthines, methylated purines, hydroxyphenyl valeric acid derivatives
Sugar sweetened soft drinks/confectionary	18	Urine	Non-essential amino acids, dipeptides, disaccharides
Artificially sweetened drinks	5	Urine	Synthetic disaccharide derivatives
Alcoholic drinks			
Alcohol general	8	Serum	Fatty acids, uronic acids
Wine	87	Urine	Phenolic acids, monocarboxylic acids, valerolactones, gallates
Beer/Lager	2	Serum and urine	Phytoestrogen, fatty acids
Fruit and vegetables			
General fruit and vegetables	32	Urine	Hippuric acids
Spices	46	Plasma	Wide range of compounds
Specific fruit			
Apple	102	Urine	Flavanols, phenylvalerolactones, dihydrochalcones, sugar alcohols, hydroxycinnamates
Pear	21	Plasma	Sugars, sugar acids
Banana	19	Plasma	Monoamine derivatives, methoxybenzoate derivatives, organic acids
Grapes	41	Urine	Anthocyanin derivatives, dicarboxylic acid, flavanols, hydroxycinnamic acids, phenylvaleric acid derivatives
Strawberries/red berries	30	Urine	Hydroxybenzoic acids, hydroxycoumarins, anthocyanins, furan, and furanone derivatives
Tomatoes	29	Plasma and Urine	Flavanol derivatives, hydroxycinnamic acids, flavanones, steroidal, and β-carboline alkaloids
Specific vegetables			
Cruciferous vegetables	24	Urine	Isothiocyanate derivatives, sulfhydryl compounds, organic acids, flavone derivatives, carboxylic acids, vitamins
Other leafy greens/spinach/lettuce	8	Plasma	Tetrapyrrole derivatives, guanidino derivatives, flavanone derivatives
Carrots	6	Plasma	Carotenoids
Onion/garlic	20	Urine	Sulfhydryl derivatives, flavanol derivatives
Mushrooms	1	Plasma and serum	Histidine derivative
Starchy foods/vegetable protein			
Legumes (peas, beans, peanuts)	54	Urine	Di- and tri-peptides, amino acid derivatives, vitamin B derivatives
Soy products	11	Urine	Isoflavonoids
Nuts	56	Urine	Flavanol derivatives, dihydroxyphenylacetic acid derivative, valerolactones,
Potatoes	5	Serum and urine	Glycoalkaloids, nortropane alkaloids
Cereal products, pasta, rice			
White bread	9	Urine	Acyl-carnitines, benzoxazines
Wholegrain general	41	Urine	Alkylresorcinols, dihydroxybenzoic acids, benzoxazinoids
Wholegrain rye	21	Urine	Benzoxazinoids, carboxylic acids, dibenzoic acids, caffeoylquinic acid derivatives
Rice	1	Serum	Long-chain fatty acid derivative
Baked or toasted products			
Strongly heated foods	11	Urine	Pyrrolines, pyridine derivatives
Baked or toasted grain products	1	Urine	Pyrroline
Fried foods	15	Urine	Furaneols, pyrrolines, furanone derivatives
Smoked foods	8	Urine	Syringol derivatives

(*Continued*)

Table 1. Continued.

Dietary components targeted in MAIN study menus	No. of putative biomarkers	Main biofluid type	Main chemical classes
Eggs and Dairy			
General dairy products	7	Plasma	Phosphatidyl cholines, non-steroidal oestrogen derivatives
Eggs	1	Serum	Unsaturated long-chain fatty acid
Butter	4	Plasma and serum	Fatty acids
Cheese and other dairy products	13	Urine	Organic acid derivatives, acyl-glycine derivatives, quinoline derivatives
Meats			
General meat	20	Urine	Carnitine derivatives, histidine derivatives, hydroxyprolines, non-protein amino acids
Red meat	20	Urine	Histidine derivatives, hydroxyprolines, dipeptides,
Processed meats (inc. cured, smoked)	4	Urine	Acylcarnitine, methylthiazolidine derivatives
Poultry	4	Urine	Histidine derivatives, dipeptides
Fish (all)	6	serum	Amines, long-chain fatty acids, furoic acid derivatives
Fish (oily, fatty)	15	Plasma	Amines, long-chain fatty acids, furoic acid derivatives, dipeptides
Fish (white)	4	Plasma and urine	Betaine derivatives, carboline derivatives
Shellfish	5	Urine	Organic acids, amines
Poultry and fish	2	Urine	Dipeptide, histidine derivative
Other meat (offal)	1	Urine	Mercapturic acid conjugate
Protein intake and digestion			
Protein intake and digestion	37	Urine	Aromatic amino acid breakdown products, glutamine conjugates

(e.g., peak shape, tailing, and resolution) and ionization conditions for selected metabolites present at the same concentration in standard mixtures showed that this combination of columns allowed separation and simultaneous quantification of the panel of biomarkers by MRM methodology (data not shown).

Figure 1 illustrates the behavior of selected food biomarker candidates in the calibration mixtures on both HILIC (at 100 µg mL^{-1}) (A and B) and RP (at 30 µg mL^{-1}) (C and D) to demonstrate some general concepts relevant to the design of a biomarker panel providing comprehensive coverage of dietary exposure. Most biomarkers chosen for validation showed good symmetry and narrow peaks with little evidence of tailing; however, there is clear evidence of overlap from the MRM traces which influences absolute quantification. Where possible, biomarker candidates with poorer chromatographic qualities could be omitted from panels and quantitation should focus on related metabolites with better performance (e.g., 3-methyl xanthine could be used in preference to 7-methyl xanthine; Figure 1C peaks 13 and 12, respectively). Differences in ionization efficiency (and hence signal intensity) between putative biomarkers is an important consideration when that metabolite is usually present at low concentrations. The use of two or more phase II metabolism products may be valuable if there is evidence of differential biotransformation between individuals due to genotypic effects (e.g., ferulic acid-4-O-β-D-glucuronide and ferulic acid-4-O-sulphate; Figure 1C peaks 15 and 17). For biomarkers expected to be present in high concentrations, it may be valuable to utilize two related molecules (especially where each is found at different concentrations) to ensure that at least one target is quantifiable within the linear range of the calibration curve (e.g., hippuric acid and 4-hydroxy hippuric acid; Figure 1C peaks 18 and 14).

3.2. Validation of an Analysis Strategy for Dietary Exposure Biomarkers on Reverse Phase and Hydrophilic Interaction Liquid Chromatography Columns

For both HILIC and RP analyses, chemical standard calibration mixtures were used to quantify biomarker concentrations for each batch of samples. The LOD and LOQ of candidate biomarkers were determined (**Table 2**). The concentrations of the full panel of biomarkers were measured in pooled ($N = 9$) urine samples from each of 23 free-living participants in Study 3 (Table 2). The median concentrations ranged across 5 orders of magnitude from >1 mg per mL for hippurate and creatinine to approaching nanogramme levels for other metabolites (e.g., 1,3-benzoxazol-2-one (BOA), chlorogenic acid, and D,L-sulforaphane L-cysteine) with the vast majority metabolites falling in the range 0.05–100 µg mL^{-1}. With the exception of quercetin and vanillic acid, the median concentrations in urine pools were above the LOQ for each biomarker.

QC measurements of calibration standards and urine QC samples (**Figure 2**A) showed a clear performance improvement when utilizing RP chromatography versus HILIC. In a standard calibration mix, 32 biomarkers measured under RP chromatography conditions were below 20% RSD, with 12 of the 32 < 10% and no biomarkers > 30%. Biomarkers measured under HILIC conditions showed similar reproducibility, but with 3 biomarkers > 30%. When reproducibility was measured in a standard urine QC sample, there were higher RSD values for some biomarkers (in total 16 out of 62 > 30%). To further quantify the reproducibility of biomarkers, RSD values were measured following repeated acquisitions of a urine QC sample over a 14 days period (Figure 2B). Although inter-day RSD values were higher, the median

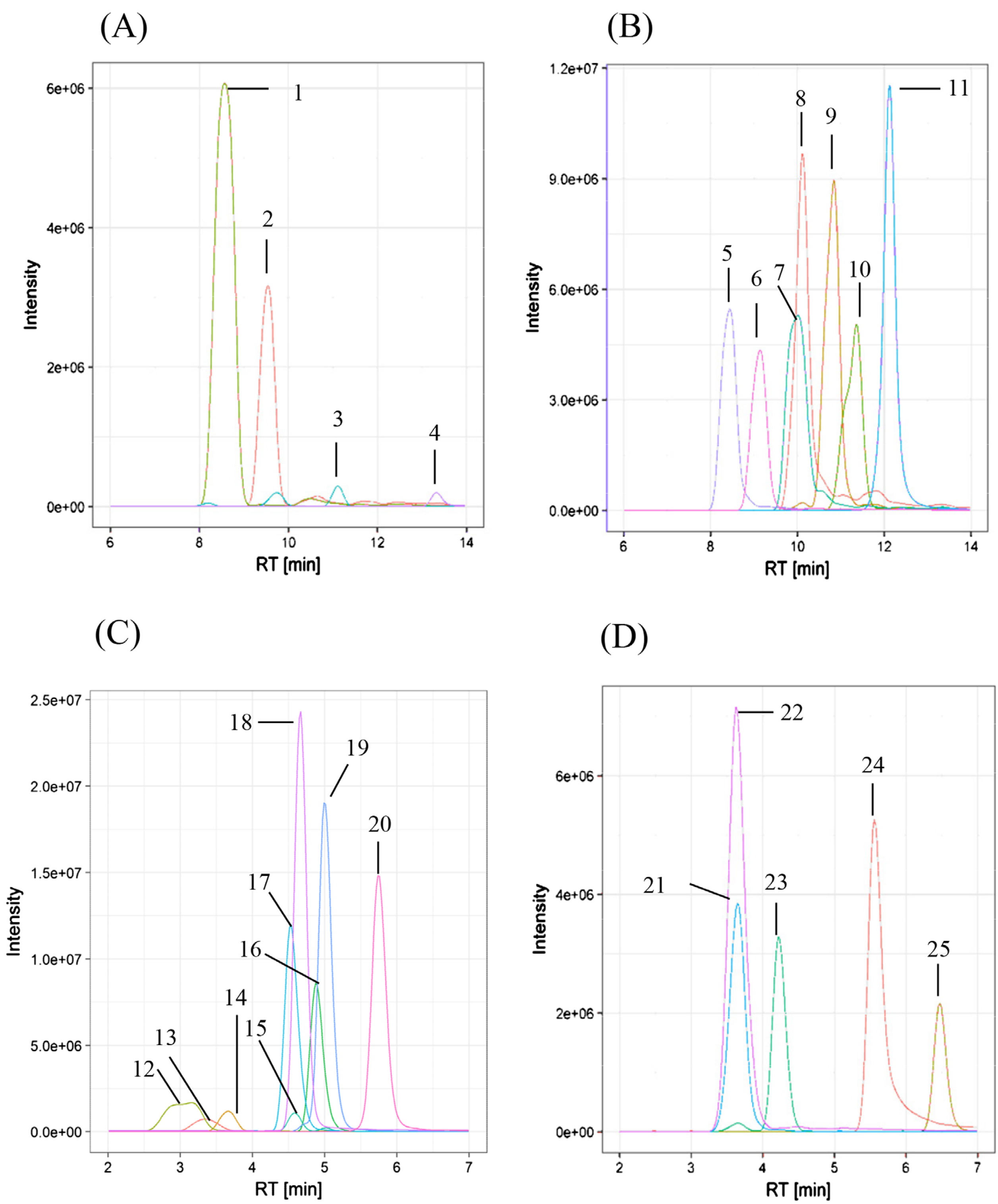


Figure 1. . Chromatographic representation of urinary biomarkers for HILIC and RP methods for selected foods and food groups A) HILIC, citrus, grape, sugary foods, and drinks. B) HILIC, red meat, poultry, white fish, and oily fish. C) RP-$C_{18,}$ fruits and vegetables, flavonoid-rich and anthocyanin rich foods, crucifers, cocoa products and coffee. D) RP-$C_{18,}$ whole grain foods, sourdough, strongly heated (baked) foods, soy. Where: (1) proline betaine, (2) 4-hydroxyproline betaine, (3) D-sucrose, (4) L-tartarate, (5) creatinine, (6) trimethylamine-*N*-oxide, (7) carnitine, (8) 1-methylhistidine, (9) 3-methylhistidine, (10) L-anserine, (11) carnosine, (12) 7-methylxanthine, (13) 3-methyl xanthine, (14) 4-hydroxyhippuric acid, (15) ferulic acid-4-O-β-D-glucuronide, (16) D,L-sulforaphane-*N*-acetyl-L-cysteine, (17) ferulic acid-4-O-sulfate, (18) hippuric acid, (19) feruloylglycine, (20) quercetin-3-O-β-D-glucuronide, (21) 3-(3,5-dihydroxyphenyl)-1-propanoic acid (DHPPA)-3-sulfate, (22) *N*-(2-furoyl)glycine, (23) DHPPA, (24) BOA (1,3-benzoxazol-2-one), (25) daidzein.

Table 2. Limit of Detection (LoD), limit of quantification (LoQ), and summary of concentration ranges of dietary biomarkers in urine collected from free-living individuals in Study 3.

Biomarker	Dietary component	Concentration		Concentration in urine		
		LoD	LoQ	Median	Min	Max
		[μg mL^{-1}]		[μg mL^{-1}]		
Hippuric acid	Fruit and vegetables	0.00059	0.00197	2380.68	1011.08	3421.68
Creatinine	Meat	0.00009	0.00030	1993.81	614.06	6291.30
Phenyl-acetyl-L-glutamine	Protein	0.00164	0.00548	124.52	13.65	338.22
Indoxyl-sulfate	Protein (colon fermented)	0.00011	0.00036	106.20	25.48	325.00
7-Methyl-xanthine	Cocoa	0.00048	0.00158	97.88	16.40	390.09
1-Methyl histidine	Meat	0.00142	0.00472	96.92	22.51	270.28
Trimethylamine-*N*-oxide	Seafood	0.00009	0.00030	78.63	0.13	356.67
3-Methyl histidine	Poultry/fish	0.00114	0.00381	60.23	4.31	223.97
Caffeine	Coffee/cocoa	0.00011	0.00036	55.67	3.79	162.31
N-(2-Furoyl)glycine	Strongly heated foods	0.00075	0.00250	49.71	11.49	105.46
Ferulic acid-4-O-sulfate	Coffee	0.00087	0.00289	47.32	13.20	116.98
Feruloylglycine	Coffee	0.00234	0.00779	46.68	9.72	218.19
p-Cresol-sulfate	Protein (colon fermented)	0.00034	0.00112	32.92	0.23	142.91
L-Histidine	Protein intake	0.00024	0.00081	27.49	6.58	102.13
Trigonelline	Legumes	0.00114	0.00381	13.02	2.82	77.53
Pyrogallol	Legumes	0.00262	0.00872	0.01	0.00	5.28
Taurine	Meat	0.00298	0.00993	11.40	0.09	68.80
Carnitine	Meat	0.00007	0.00024	10.78	0.83	62.09
Acesulfame-K	Sweetener	0.00013	0.00042	10.16	0.00	694.74
3-Hydroxyhippuric acid	Fruit and vegetables	0.00059	0.00197	9.59	2.14	48.72
4-Hydroxyhippuric acid	Fruit and vegetables	0.00059	0.00197	9.00	3.28	54.20
Caffeic acid	Cocoa/tea	0.00133	0.00443	8.29	2.39	15.65
p-Cresol-glucuronide	Protein (colon)	0.00017	0.00055	8.12	0.07	67.87
Proline betaine	Citrus	0.00114	0.00381	7.25	0.07	89.79
Tartarate	Grapes	0.00087	0.00289	5.94	0.04	60.48
DHBA-3-O-sulfate	Whole grain	0.00059	0.00197	4.71	0.67	52.38
3-Methyl-xanthine	Cocoa	0.00234	0.00779	3.87	0.52	18.38
Quercetin-3-O-b-D-glucuronide	Onion	0.00169	0.00564	3.35	0.15	31.19
L-Tryptophan	Protein Intake	0.00006	0.00018	2.48	0.47	10.60
Sucrose	Sugary foods and drinks	0.00044	0.00146	2.45	0.01	60.80
L-Phenylalanine	Protein Intake	0.00009	0.00030	2.20	0.18	7.39
L-Anserine	Chicken	0.00028	0.00093	2.04	0.21	19.74
Rhamnitol	Apple	0.00046	0.00152	1.77	0.01	0.54
Carnosine	Meat	0.00009	0.00030	1.39	0.52	6.20
Ferulic acid	Polyphenol rich foods	0.00169	0.00564	1.28	0.19	4.06
Dopamine-4-O-sulfate	Banana	0.01181	0.03957	1.11	0.58	6.02
4-Hydroxyproline-betaine	Citrus	0.00048	0.00158	1.24	0.05	18.70
Ferulic acid-4-O-b-D-glucuronide	Polyphenol rich foods	0.00206	0.00687	1.09	0.29	6.02
Furaneol	Red berries	0.00289	0.00964	0.28	0.13	2.02
DHBA	Wholegrain	0.00142	0.00472	1.08	0.16	2.49
DHPPA	Wholegrain	0.00059	0.00197	1.06	0.29	2.91
D,L-Sulforaphane-*N*-acetyl-L-cysteine	Cruciferous Vegetables	0.00059	0.00197	0.97	0.01	30.40
DHPPA-3-sulfate	Whole grain	0.00087	0.00289	0.85	0.21	3.65
Calystegine A_3	Potatoes	0.00142	0.00472	0.55	0.03	2.40
Ethyl-beta-D-glucuronide	Alcohol	0.00014	0.00047	0.43	0.06	35.83
Dopamine-3-O-sulfate	Banana	0.01181	0.03957	0.04	0.00	0.29

(*Continued*)

Table 2. Continued.

Biomarker	Dietary component	Concentration		Concentration in urine		
		LoD	LoQ	Median	Min	Max
		[μg mL^{-1}]		[μg mL^{-1}]		
b-Alanine	Meat	0.00065	0.00217	0.37	0.03	1.66
Calystegine B_2/B_1	Potatoes	0.00011	0.00036	0.35	0.01	1.16
Dihydrocaffeic acid	Cocoa	0.00059	0.00197	0.17	0.02	2.78
Gallic-acid	Fruit/grapes/tea	0.00234	0.00779	0.15	0.03	1.20
Resveratrol	Wine	0.00048	0.00158	0.12	0.25	7.16
Chlorogenic acid	Coffee	0.00222	0.00741	0.10	0.00	0.89
D,L-Sulforaphane L-cysteine	Cruciferous vegetables	0.00024	0.00081	0.08	0.00	3.10
Epicatechin(-)	Polyphenol rich foods	0.00257	0.00858	0.03	0.00	1.51
p-Coumaric acid	Grapes/berries	0.00006	0.00018	0.03	0.01	0.41
Protocatechuic acid	Anthocyanin rich foods	0.00059	0.00197	0.03	0.03	1.57
m-Coumaric acid	Coffee	0.00011	0.00036	0.02	0.00	0.32
Daidzein	Soy	0.00114	0.00381	0.02	0.00	4.31
Naringenin	Citrus	0.00009	0.00030	0.01	0.00	0.33
BOA (1,3-Benzoxazol-2-one)	Wholegrain/rye	0.00011	0.00036	0.01	0.00	0.04
Quercetin	Onion	0.00142	0.00472	0.00	0.00	0.03
Vanillic acid	Cocoa	0.00242	0.00806	0.00	0.00	0.01

value was 19.95% and most values were below the arbitrary 30% RSD threshold often adopted in the literature.[27]

3.3. Use of Biomarker Panel to Detect Exposure to Targeted Foods in a Free-Living Study Cohort

To explore the utility of a quantitative biomarker panel to characterize eating habits within populations, we measured the concentrations of dietary exposure biomarkers in FMV urine samples obtained the day after following consumption of 3 distinctive meal plans (see Table S3A, Supporting Information, for menu details) in a food intervention study (Study 2).[15,16] The data were initially subjected to multivariate analysis using RF decision trees.[28] **Figure 3**A shows a MDS of RF proximity values for multinomial classification of liquid chromatography triple quadrupole mass spectrometry (LC-QQQ-MS) urinary panel biomarker concentrations after the three menu days. These data demonstrate distinctive groupings of urine samples representing each daily intervention diet as evidenced by the ROC curves in Figure 3B illustrating 100% specificity and sensitivity.

The average concentrations in FMV urines of individual (or combinations of) biomarkers putatively indicative of exposure to specific foods/food groups within the three menu days are shown in **Figure 4**. The relative levels of biomarkers correspond well with consumption of foods targeted on each menu day. For example, in relation to meat, 3-methylhistidine is strongly indicative of eating chicken on menu day 1, TMNO reports fish/shellfish exposure on menu day 2, whilst carnosine levels associated with red meat consumption are highest following consumption of a 100% beef burger on menu day 3. The relative exposure to overall protein in the diet is correlated well with the total concentration of aromatic amino acids in FMV urine. Exposure to multiple potato products on menu days 2 and 3 can be monitored by total calystegine content; in this case biomarker signals above baseline levels are also visible on menu day 1 due to carry over from foods consumed the previous day, depending on the randomization of the order of exposure to each menu day (data not shown). A similar pattern is evident with other foods commonly used as constituents of complex meals, including tomato and onion (quercetin), processed meats (carnitine), and baked/roasted products (N-2-furoyl-glycine). Several biomarkers show a very strong specificity to individual foods/food groups that are eaten less frequently, including grape products (tartrate), strawberries (furaneol), cocoa products (3- and 7-methyl xanthine), brassicas (D-L-sulforaphane-*N*-acetyl-L-cysteine), and artificially-sweetened drinks (acesulfame-K). Several biomarkers show some overlap between target foods. For example, pyrogallol is a biomarker of general legume dietary exposure but the presence of daidzein in the same urine samples indicate that soya products have been consumed specifically on menu day 3. Similarly, the levels of DHPPA and DHPPA sulfate on menu days 1 and 3 correlates with exposure to whole grain foods, but elevation also of BOA on the latter day suggest that rye-containing food products specifically have been eaten.

3.4. Use of Biomarkers to Quantify Dietary Exposure

We have reported previously,[18] an inpatient, randomized, controlled, crossover food intervention (Study 1) in which 19 participants consumed experimental diets that met 25%, 50%, 75%, or 100% of the dietary intake recommended by the WHO healthy eating guidelines through systematic differences in intakes of

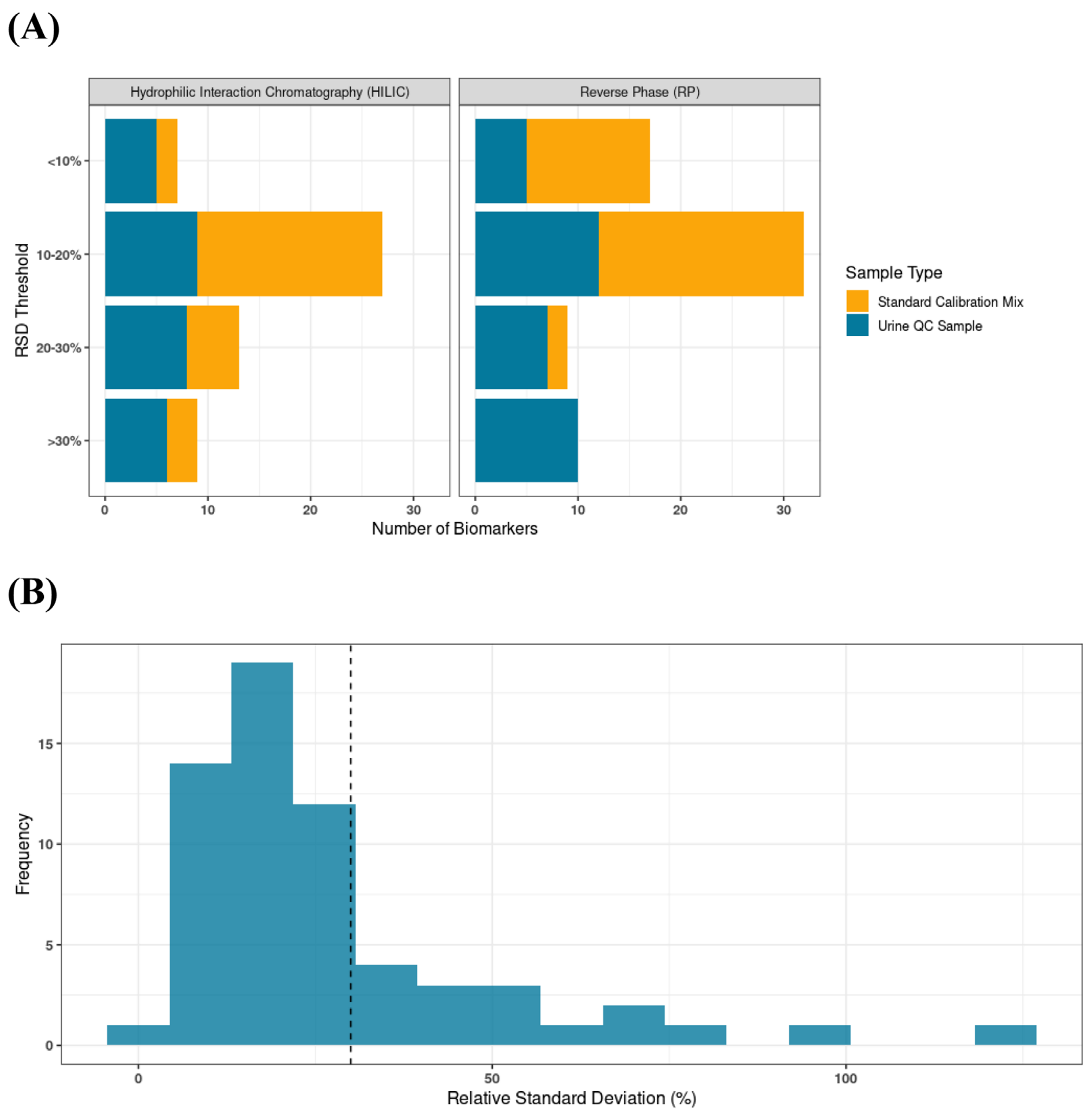


Figure 2. Reproducibility of biomarker measurement. A) Summary of relative standard deviation (RSD) thresholds from reproducibility testing of biomarkers in a calibration standard mixture and urine quality control samples derived from Study 3. B) Distribution of RSD values for external urine QC samples measured repeatedly over 14 days. Dashed red line indicates the median value (19.95%) and the black dashed line is the arbitrary 30% threshold. Four biomarkers were omitted from the RSD analysis as in this specific population pooled sample the concentrations were < LoQ.

fruits, vegetables, and whole grain foods. For example, grape products, cocoa-containing products, whole grain foods, and cruciferous vegetables were consumed at four different exposure levels (Table S3B, Supporting Information). When measured by UHPLC coupled to a QQQ-MS instrument, the concentrations in post-dinner spot urine samples of biomarkers selected to quantify exposure to each of the foods/food groups (L-tartarate, 7-methyl xanthine, DHPPA-3-sulfate and D,L-sulforaphane-*N*-acetyl-L-cysteine, respectively) correlated well with the level of consumption (**Figure 5**).

4. Discussion

A consensus-based procedure for systematic validation of biomarkers of dietary intake has been proposed recently. This procedure stresses the importance of evaluating key criteria

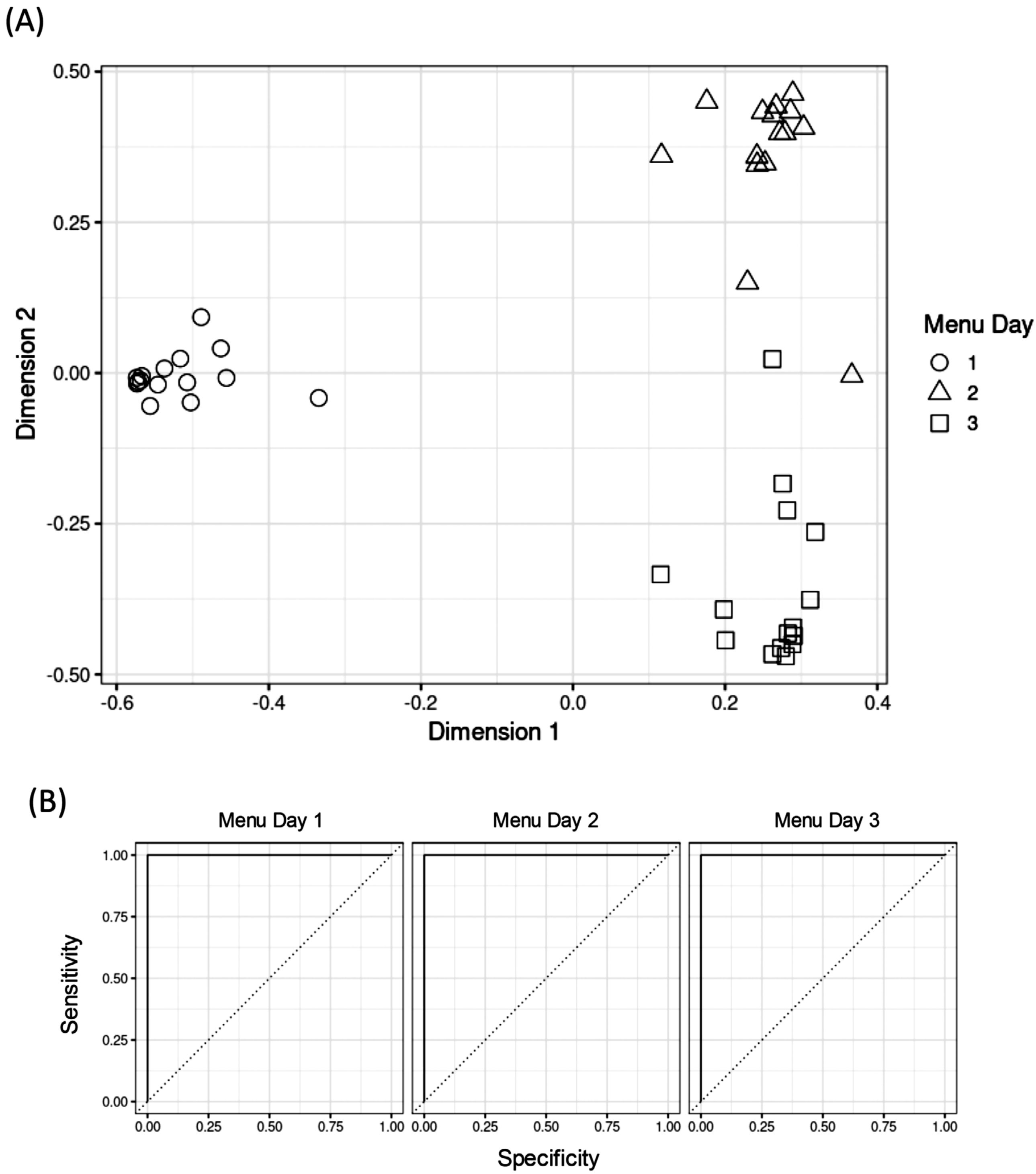
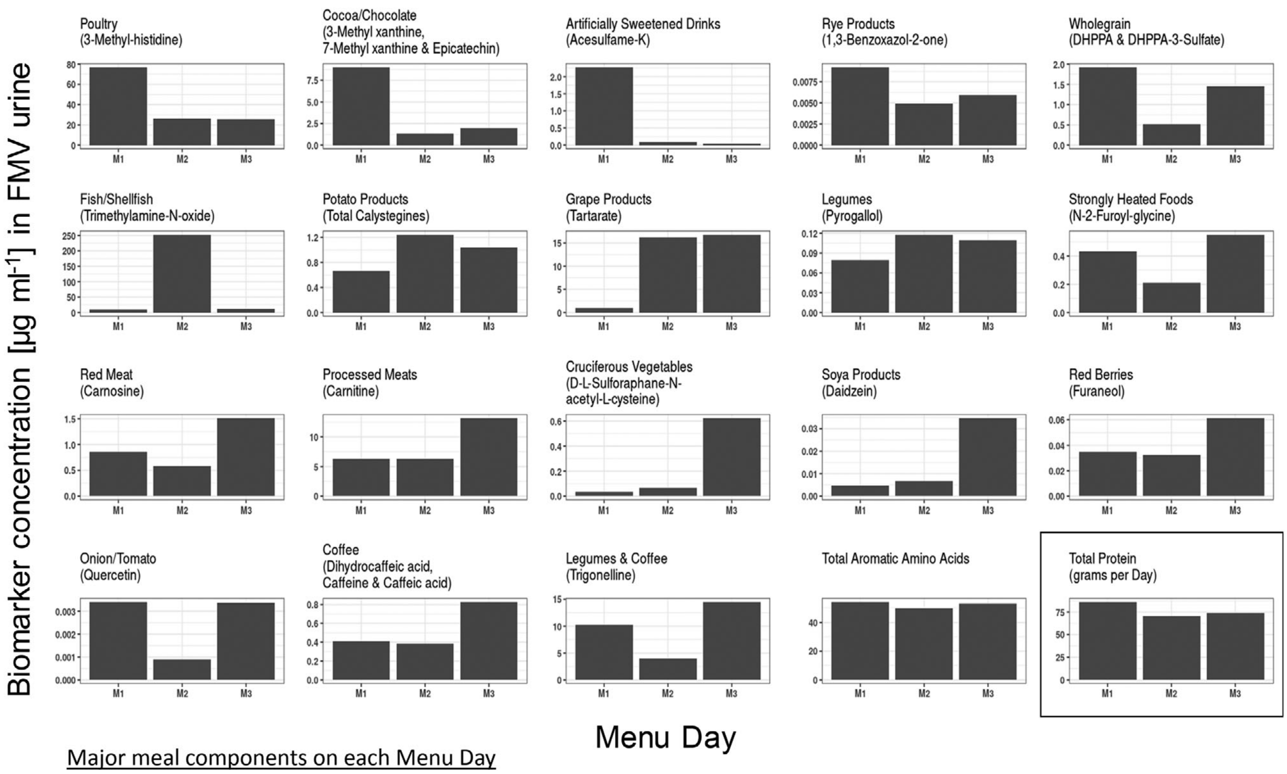


Figure 3. Detection of differential daily dietary exposure using a panel of urine biomarkers. A) Multi-dimensional scaling (MDS) of random forest proximity extracted from a classification model of 15 individuals consuming experimental diets in 3 separate treatment periods, and a selected panel of 58 dietary biomarkers. Each symbol represents the urine metabolome of an individual participant on a specific menu day. B) Individual receiver operator characteristic (ROC) curves illustrating classification performance to discriminate urine samples from each menu day.



Major meal components on each Menu Day

Menu Day 1 (M1): [Breakfast] Wholegrain bread (rye) & Baked beans, Sugar puffs & semi-skimmed milk; [Lunch] Chicken breast, Tomatoes, Cucumber, Wholegrain bread (rye), Hot chocolate; [Dinner] Wholegrain pasta & Beef Bolognese, Garlic bread, Chocolate bar, Pear, Diet soft drink

Menu Day 2 (M2): [Breakfast] Tea, Red grape juice, Egg, Potato waffles, White bread; [Lunch] Tea, Prawn sandwich (white bread), Red Grapes; [Dinner] Cod fish fingers, Jacket potato, Cauliflower, Mushy peas, Peanuts, Sparkling red grape juice

Menu Day 3 (M3): [Breakfast]Coffee, White grape juice, Pork sausages, Wholegrain bread roll, Cornflakes & Soya milk; [Lunch] Coffee, Processed cheese, Wholegrain bread roll, Sausage roll, Red pepper, White grapes; [Dinner] 100% Beef burger, Oven chips, Coleslaw, Strawberries, Yogurt

Figure 4. Dietary exposure biomarker concentrations in FMV urine following consumption of 3 distinctive menu plans. Bar charts show average biomarker concentration (µg mL^{-1}) in normalized FMV urine samples (n = 15) after exposure to menu plans (Table S3A, Supporting Information: M1 = menu day 1, M2 = menu day 2, M3 = menu day 3). Major meal components consumed on each menu day are summarized at the bottom of the figure.

including dose- and time-responsiveness, robustness, analytical performance, and inter-laboratory reproducibility.[29] Although exhaustive validation of individual biomarkers for specific foods is clearly important, because the whole dietary pattern may influence health outcomes,[30,31] there is also considerable interest in evaluating overall eating habits within populations.[32,33] This demands development of a comprehensive panel of biomarkers. As a step toward this objective using urine-based dietary intake biomarkers, we have demonstrated that the chromatographic resolution and sensitivity provided by 2 standardized MRM LC-QQQ-MS/MS methods is sufficient to enable routine separation and quantification of complex mixtures of dietary-related metabolites spanning a wide range of polarities and several orders of magnitude in concentration. Observations made during the present study allow determination of biomarker attributes that impact on decisions concerning their eventual inclusion in a validated panel for routine use in the future.[29]

For a dietary exposure biomarker panel to have real-world utility, coverage should include all those foods that are consumed frequently and/or in large amounts by the population under study plus any key foods and food groups that have particular (health) significance.[11,32,34,35] This ambition has to be balanced by the need to be realistic about those eating behavior factors that influence the feasibility of obtaining accurate measurement. This will be a particular constraint for foods that are consumed in small amounts and/or infrequently and when using one, or few, urine samples per participant.[36] Importantly, since very few metabolite markers are unique to individual foods, future validation studies will need to be able to examine simultaneously the behaviors of multiple biomarker candidates in the context of a diverse a range of dietary intake. As a step toward this objective, we have described the selection of target foods/food groups and the development and implementation of food intervention study designs to represent the UK diet,[15,16] together with the development of urine sampling methods, suitable for large scale use in community settings.[37,38] Using urine samples and dietary exposure records from these intervention studies, in the present study we have demonstrated that simultaneous, quantitative monitoring of several biomarkers already known to have high specificity for foods commonly consumed in the UK (including whole grain products, cruciferous vegetables, cocoa, and grape products) is feasible using LC-MS/MS. In circumstance where biomarkers

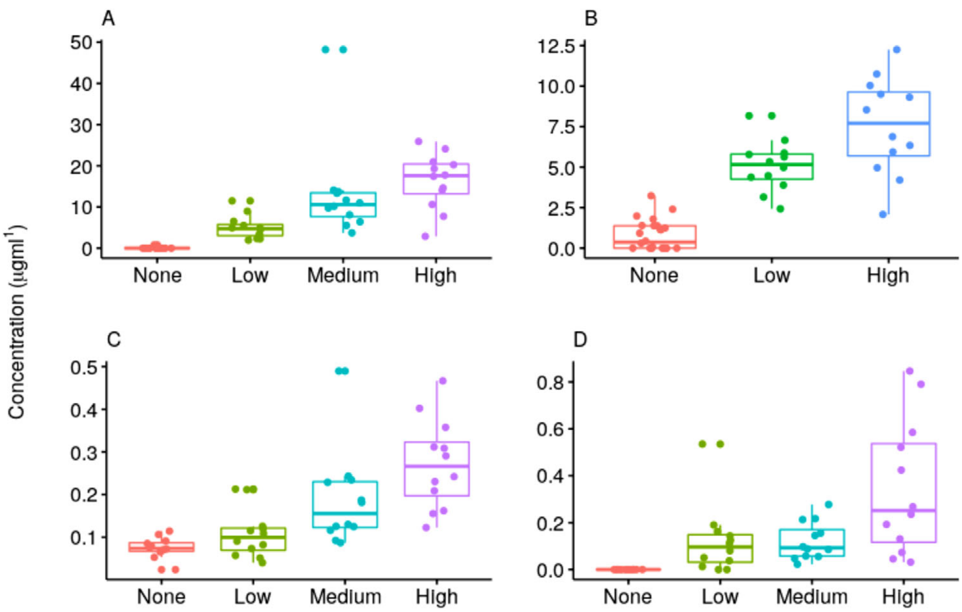


Figure 5. Boxplots of urinary concentrations of food intake biomarkers. Where, A) L-tartarate, B) 7-methyl xanthine, C) 3-(3,5-dihydroxyphenyl)-1-propanoic acid (DHPPA-3-sulfate), D) D-L-sulforaphane-*N*-acetyl-L-cysteine) following graded intakes of specific foods under controlled conditions from Study 1. Menu plans for days representing none, low, medium, and high exposure to the target foods are shown in Table S3B, Supporting Information.

are associated with exposure to more than one food group (e.g., trigonelline reports exposure to both legumes and coffee) then understanding signal ratios in relation to other putative biomarkers will be important during future validation as described recently for discrimination of meat exposure[39]

The routine implementation of a comprehensive dietary exposure monitoring tool requires control of the major sources of variability, both biological and technical. From a biological perspective, some dietary chemicals undergo different metabolic fates as a result of inter-individual differences in microbiome/endogenous metabotype within populations. This problem may be overcome by measuring an appropriate combination of signals (metabolites) as highlighted in recent reports.[40–43] For example, the chemical complexity of the urine metabolome caused by endogenous glucuronidation and sulfation activities is well documented.[44,45] It is also important to address the variability associated with human micturition behavior. In several studies, we have shown that analysis of FMV spot urine samples is highly informative of dietary exposure and that, as well as being much easier to collect and transport, FMV urine is an adequate substitute for 24 h urine collection.[16,36–38] Normalization of the concentration of individual spot urine samples by addition of an appropriate amount of water, guided by simple refractometry measurements, has proved to be effective in facilitating routine analysis of large numbers of samples.[19]

In the present study we have described quantitative approaches using QQQ-MS to measure biomarker abundance which are dependent on the use of complex mixtures of chemicals standards for calibration. Others have shown that similar approaches for "quantitative dietary fingerprinting," particularly for polyphenol-derived metabolites, are feasible using less sensitive hybrid quadrupole/ion trap technology when combined with urine concentration by solid phase extraction methodology.[43] Commercial availability, purity, stability, solubility, and cost of a chemical standard are important considerations in the adoption of any biomarker for inclusion in a standardized biomarker panel. In the present investigation, intra- and inter-day reproducibility of quantification of QC samples was excellent. However, it was more challenging to achieve the same levels of accuracy and precision with real-world urine samples because constituents targeted for quantification exhibited unexpected ionization behaviors on occasion, potentially due to interference by other urine constituents. However, future validation studies should examine the behavior of alternative putative biomarkers in relation to expected concentration ranges within a population and should consider the frequency of unexpected ionization behavior caused by coeluting compounds when making final selections. We suggest that the analysis of the behavior of putative biomarkers in a "master mix" of urine samples representative of the study population is a valuable step in the selection process.

In conclusion, the analytical methodology outlined in the present study provides a useful framework for developing and validating panels of comprehensive biomarkers with utility in real-world assessment of dietary exposure within free-living populations. Such approaches could be used to validate, and to complement, information derived from traditional, self-reported dietary assessment instruments. This will help to mitigate the pervasive, and difficult to quantify, problems of participant misreporting, bias, and inaccuracies of dietary recording and so improve future nutrition studies.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

M.B. and T.W. contributed equally to this work. The authors thank the volunteers for their commitment and the Clinical Research Facilities in Newcastle, London, and Aberystwyth for volunteer support. The authors would like to thank Elaine Holmes, who helped to coordinate research and Dr. Edward Chambers, who coordinated volunteer CRF visits and supervised support staff in Imperial College London (Study 1) and Dr. Naomi Willis and Dr. Long Xie, who coordinated volunteer CRF visits and supervised support staff in Newcastle (Study 2). This work was funded by MRC grant ref: MR/ J010308/1.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Authors' contributions to the manuscript are as follows: M.B., designed QQQ experiments, assessed chromatographic behavior characteristics, supervised MS support staff, and wrote the manuscript; T.W., assessed chromatographic behavior characteristics, QQQ data processing, classification modelling, produced figures, and wrote the manuscript; H.Z., QQQ technical support, selected reaction monitoring parameters, and data generation; A.J.L., researched biomarker literature and identified putative urinary biomarkers; K.T., QQQ analysis and technical support in Aberystwyth; G.F. and I.G.-P. supervised Study 1 in Imperial College London; J.C.M. supervised Study 2 in Newcastle and edited the manuscript; R.T. supervised Study 3 in Aberystwyth; N.G. carried out Study 3 RCT; J.D. coordinated overall project, supervised research in Aberystwyth, developed the biomarker panel strategy, designed figures, and wrote the manuscript. All authors read and approved the final manuscript.

Keywords

dietary biomarkers, liquid chromatography mass spectrometry, metabolomics, targeted quantification, urine

Received: June 2, 2020
Published online:

[1] L. O. Dragsted, *Am. J. Clin. Nutr.* **2017**, *105*, 8.
[2] M. Beckmann, A. M. Joosen, M. M. Clarke, O. Mugridge, G. Frost, B. Engel, K. Taillart, A. J. Lloyd, J. Draper, J. K. Lodge, *Mol. Nutr. Food Res.* **2016**, *60*, 444.
[3] H. Gibbons, B. A. McNulty, A. P. Nugent, J. Walton, A. Flynn, M. J. Gibney, L. Brennan, *Am. J. Clin. Nutr.* **2015**, *101*, 471.
[4] A. J. Lloyd, G. Fave, M. Beckmann, W. Lin, K. Tailliart, L. Xie, J. C. Mathers, J. Draper, *Am. J. Clin. Nutr.* **2011**, *94*, 981.
[5] M. C. Playdon, S. C. Moore, A. Derkach, J. Reedy, A. F. Subar, J. N. Sampson, D. Albanes, F. Gu, J. Kontto, C. Lassale, L. M. Liao, S. Männistö, A. M. Mondul, S. J. Weinstein, M. L. Irwin, S. T. Mayne, R. Stolzenberg-Solomon, *Am. J. Clin. Nutr.* **2017**, *105*, 450.
[6] M. Kristensen, S. B. Engelsen, L. O. Dragsted, *Metabolomics* **2012**, *8*, 64.
[7] G. Praticò, Q. Gao, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, L. Brennan, S. H. Pedapati, L. A. Afman, D. S. Wishart, R. Vázquez-Fresno, C. Andres-Lacueva, M. Garcia-Aloy, H. Verhagen, E. J. M. Feskens, L. O. Dragsted, *Genes Nutr.* **2018**, *13*, 3.
[8] M. Garcia-Aloy, M. Rabassa, P. Casas-Agustench, N. Hidalgo-Liberona, R. Llorach, C. Andres-Lacueva, *Trends Food Sci. Technol.* **2017**, *69*, 220.
[9] A. Scalbert, L. Brennan, C. Manach, C. Andres-Lacueva, L. O. Dragsted, J. Draper, S. M. Rappaport, J. J. J. van der Hooft, D. S. Wishart, *Am. J. Clin. Nutr.* **2014**, *99*, 1286.
[10] S. Bouatra, F. Aziat, R. Mandal, A. C. Guo, M. R. Wilson, C. Knox, T. C. Bjorndahl, R. Krishnamurthy, F. Saleem, P. Liu, Z. T. Dame, J. Poelzer, J. Huynh, F. S. Yallou, N. Psychogios, E. Dong, R. Bogumil, C. Roehring, D. S. Wishart, *PLoS One* **2013**, *8*, e73076.
[11] M. C. Playdon, J. N. Sampson, A. J. Cross, R. Sinha, K. A. Guertin, K. A. Moy, N. Rothman, M. L. Irwin, S. T. Mayne, R. Stolzenberg-Solomon, S. C. Moore, *Am. J. Clin. Nutr.* **2016**, *104*, 776.
[12] N. R. Kitteringham, R. E. Jenkins, C. S. Lane, V. L. Elliott, B. K. Park, *J. Chromatogr. B* **2009**, *877*, 1229.
[13] R. Wei, G. Li, A. B. Seymour, *Anal. Chem.* **2010**, *82*, 5527.
[14] M. Beckmann, A. J. Lloyd, S. Haldar, C. Seal, K. Brandt, J. Draper, *Mol. Nutr. Food Res.* **2013**, *57*, 1859.
[15] A. J. Lloyd, N. D. Willis, T. Wilson, H. Zubair, E. Chambers, I. Garcia-Perez, L. Xie, K. Tailliart, M. Beckmann, J. C. Mathers, J. Draper, *Metabolomics* **2019**, *15*, 72.
[16] A. J. Lloyd, N. D. Willis, T. Wilson, H. Zubair, L. Xie, E. Chambers, I. Garcia-Perez, K. Tailliart, M. Beckmann, J. C. Mathers, J. Draper, *Mol. Nutr. Food Res.* **2019**, *0*, 1900062.
[17] World Health Organization, Geneva **2004**, ISBN 92 4 159222 2, pp. 1–18.
[18] I. Garcia-Perez, J. M. Posma, R. Gibson, E. S. Chambers, H. Hansen, H. Vestergaard, T. Hansen, M. Beckmann, O. Pedersen, P. Elliott, J. Stamler, J. K. Nicholson, J. Draper, J. C. Mathers, E. Holmes, G. Frost, *Lancet Diabetes Endocrinol.* **2017**, *5*, 184.
[19] W. M. B. Edmands, P. Ferrari, A. Scalbert, *Anal. Chem.* **2014**, *86*, 10925.
[20] L. Martens, M. Chambers, M. Sturm, D. Kessner, F. Levander, J. Shofstahl, W. H. Tang, A. Römpp, S. Neumann, A. D. Pizarro, L. Montecchi-Palazzi, N. Tasman, M. Coleman, F. Reisinger, P. Souda, H. Hermjakob, P.-A. Binz, E. W. Deutsch, *Mol. Cell. Proteomics* **2011**, *10*, R110.000133.
[21] M. C. Chambers, B. Maclean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T. A. Baker, M.-Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S. L. Seymour, L. M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E. W. Deutsch, R. L. Moritz, J. E. Katz, D. B. Agus, M. MacCoss, D. L. Tabb, P. Mallick, *Nat. Biotechnol.* **2012**, *30*, 918.
[22] R Core Team, *R Foundation for Statistical Computing*, Vienna, Austria **2013**.
[23] A. Liaw, M. Wiener, *R News* **2002**, *2*, 18.
[24] D. P. Enot, W. Lin, M. Beckmann, D. Parker, D. P. Overy, J. Draper, *Nat. Protoc.* **2008**, *3*, 446.
[25] E. Konukoglu, M. Ganz, *arXiv* **2014**, *1410.2838*, 1.
[26] B. Buszewski, S. Noga, *Anal. Bioanal. Chem.* **2012**, *402*, 231.
[27] O. Fiehn, J. Kopka, P. Dörmann, T. Altmann, R. N. Trethewey, L. Willmitzer, *Nat. Biotechnol.* **2000**, *18*, 1157.
[28] L. Breiman, *Mach. Learn.* **2001**, *45*, 5.
[29] L. O. Dragsted, Q. Gao, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, L. Brennan, L. A. Afman, D. S. Wishart, C. A. Lacueva, M. Garcia-Aloy, H. Verhagen, E. J. M. Feskens, G. Praticò, *Genes Nutr.* **2018**, *13*, 14.
[30] A. Granic, K. Davies, A. Adamson, T. Kirkwood, T. R. Hill, M. Siervo, J. C. Mathers, C. Jagger, *J. Nutr.* **2016**, *146*, 265.
[31] O. M. Shannon, B. C. M. Stephan, A. Granic, M. Lentjes, S. Hayat, A. Mulligan, C. Brayne, K.-T. Khaw, R. Bundy, S. Aldred, M. Hornberger, S.-M. Paddick, G. Muniz-Tererra, A.-M. Minihane, J. C. Mathers, M. Siervo, *Am. J. Clin. Nutr.* **2019**, *110*, 938.

[32] C. Galbete, J. Kröger, F. Jannasch, K. Iqbal, L. Schwingshackl, C. Schwedhelm, C. Weikert, H. Boeing, M. B. Schulze, *BMC Med.* **2018**, *16*, 99.
[33] P. F. Jacques, K. L. Tucker, *Am. J. Clin. Nutr.* **2001**, *73*, 1.
[34] W. Cheung, P. Keski-Rahkonen, N. Assi, P. Ferrari, H. Freisling, S. Rinaldi, N. Slimani, R. Zamora-Ros, M. Rundle, G. Frost, H. Gibbons, E. Carr, L. Brennan, A. J. Cross, V. Pala, S. Panico, C. Sacerdote, D. Palli, R. Tumino, T. Kühn, R. Kaaks, H. Boeing, A. Floegel, F. Mancini, M.-C. Boutron-Ruault, L. Baglietto, A. Trichopoulou, A. Naska, P. Orfanos, A. Scalbert, *Am. J. Clin. Nutr.* **2017**, *105*, 600.
[35] A. J. Lloyd, M. Beckmann, T. Wilson, K. Tailliart, D. Allaway, J. Draper, *Metabolomics* **2017**, *13*, 15.
[36] A. J. Lloyd, M. Beckmann, S. Haldar, C. Seal, K. Brandt, J. Draper, *Am. J. Clin. Nutr.* **2013**, *97*, 377.
[37] T. Wilson, N. D. Willis, H. Zubair, M. Beckmann, L. Xie, K. Tailliart, A. J. Lloyd, J. C. Mathers, J. Draper, *Proc. Nutr. Soc.* **2016**, *75*, E240.
[38] T. Wilson, I. Garcia-Perez, J. M. Posma, A. Lloyd, E. S. Chambers, K. Tailliart, H. Zubair, M. Beckmann, J. C. Mathers, E. Holmes, G. Frost, J. Draper, *J. Nutr.* **2019**, *149*, 1692.
[39] C. Cuparencu, Å. Rinnan, M. P. Silvestre, S. D. Poppitt, A. Raben, L. O. Dragsted, M. P. Silvestre, S. D. Poppitt, *Eur. J. Nutr.* **2020**.
[40] T. A. Clayton, D. Baker, J. C. Lindon, J. R. Everett, J. K. Nicholson, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 14728.
[41] C. H. Johnson, J. Ivanisevic, G. Siuzdak, *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 451.
[42] M. G. Rooks, W. S. Garrett, *Nat. Rev. Immunol.* **2016**, *16*, 341.
[43] R. González-Domínguez, M. Urpi-Sarda, O. Jáuregui, P. W. Needs, P. A. Kroon, C. Andrés-Lacueva, *J. Agric. Food Chem.* **2019**, *68*, 1851.
[44] A. R. Rechner, M. A. Smith, G. Kuhnle, G. R. Gibson, E. S. Debnam, S. K. S. Srai, K. P. Moore, C. A. Rice-Evans, *Free Radical Biol. Med.* **2004**, *36*, 212.
[45] A. Stalmach, W. Mullen, D. Barron, K. Uchida, T. Yokota, C. Cavin, H. Steiling, G. Williamson, A. Crozier, *Drug. Metab. Dispos.* **2009**, *37*, 1749.